

ABSTRACT

A method and an apparatus for processing a document of a tagged internal structure made up of a plurality of elements. A plurality of documents received by a receiving unit are stored in a RAM of a storage unit provided in a main body portion of the apparatus. The characteristic information representing the characteristics of a document is extracted in accordance with the sequence of operations recorded on a ROM under control by controller. Each document is classified into classification items making up a classification model, depending on the degree of interrelation between the characteristic information of the document extracted by a characteristic information extraction unit and the classification- item-based information to realize automatic document classification in such a manner as to reflect the interest of a user.

COPY-REPRODUCTION